4. Joint Probability Distributions

the preceding sections: one-dimensional sample spaces and a single random variable

4. Joint Probability Distributions

the preceding sections: one-dimensional sample spaces and a single random variable

situations: desirable to record the simultaneous outcomes of several random variables.

4. Joint Probability Distributions

the preceding sections: one-dimensional sample spaces and a single random variable

situations: desirable to record the simultaneous outcomes of several random variables.

Examples

1.we might measure the amount of precipitate P and volume V of gas released from a controlled chemical experiment; we get a two-dimensional sample space consisting of the outcomes (p, v).

Joint Probability Distribution

2. In a study to determine the **likelihood of success in college**, based on high school data, one might use a **three-dimensional** sample space and record for each individual his or her **aptitude test score**, **high school rank** in class, and **grade-point average** at the end of the freshman year in college.

Joint Probability Distribution

X and Y are two discrete random variables, the **joint probability distribution** of X and Y is

$$p(x, y) = P(X = x, Y = y)$$

the values $p(x, y)$ give the probability that outcomes x and y occur at the same time.

Joint Probability Distribution

Definition 3.8 The function $p(x, y)$ is a **joint probability distribution** or **probability mass function** of the discrete random variables X and Y if

1. $p(x, y) \geq 0$, for all (x, y),

2. $\sum_x \sum_y p(x, y) = 1$,

3. $P(X = x, Y = y) = p(x, y)$.

For any region A in the xy plane,

$$P[(X, Y) \in A] = \sum \sum_A p(x, y).$$

Joint Probability Distribution

Example 3.8

Two refills for a ballpoint pen are selected at random from a box that contains 3 blue refills, 2 red refills, and 3 green refills. If X is the number of blue refills and Y is the number of red refills selected, find

(a) the joint probability function $p(x, y)$

(b) $P[(X, Y) \in A]$ where A is the region $\{(x, y) | x + y \leq 1\}$.

Joint Probability Distribution

Solution

the possible pairs of values (x, y) are $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(0, 2)$, and $(2, 0)$.

Joint Probability Distribution

Solution

the possible pairs of values (x, y) are $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(0, 2)$, and $(2, 0)$.

$p(x, y)$ represents the probability that **x blue** and **y red** refills are selected.

Joint Probability Distribution

Solution

the possible pairs of values (x, y) are $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(0, 2)$, and $(2, 0)$.

$p(x, y)$ represents the probability that **x blue** and **y red** refills are selected. The total number of equally likely ways of selecting any 2 refills from the 8 is C_8^2.

Joint Probability Distribution

Solution

the possible pairs of values (x, y) are $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(0, 2)$, and $(2, 0)$.

$p(x, y)$ represents the probability that **x blue** and **y red** refills are selected. The total number of equally likely ways of selecting any 2 refills from the 8 is C_8^2.

$$p(x, y) = \frac{C_3^x C_2^y C_3^{2-x-y}}{C_8^2}$$

for $x = 0, 1, 2; y = 0, 1, 2; 0 \leq x + y \leq 2$.

Joint Probability Distribution

Solution

the possible pairs of values (x, y) are $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(0, 2)$, and $(2, 0)$.

$p(x, y)$ represents the probability that **x blue** and **y red** refills are selected. The total number of equally likely ways of selecting any 2 refills from the 8 is C_8^2.

$$p(x, y) = \frac{C_3^x C_2^y C_3^{2-x-y}}{C_8^2}$$

for $x = 0, 1, 2; y = 0, 1, 2; 0 \leq x + y \leq 2.$

Joint Probability Distribution

present the results in **Table 3.1**

TABLE 3.1 Joint Probability Distribution for Example 3.8

$f(x, y)$		$x = 0$	$x = 1$	$x = 2$	Row totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Joint Probability Distribution

(b) $P[(X, Y) \in A] = 9/14$

Joint Probability Distribution

the continuous case

Definition 3.9 The function $f(x, y)$ is a **joint density function** of the continuous random variables X and Y if

1. $f(x, y) \geq 0$, for all (x, y),

2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$,

3. $P[(X, Y) \in A] = \int \int_A f(x, y) dx dy$

for any region A in the xy plane.

Joint Probability Distribution

When X and Y are continuous random variables, the **joint density function** $f(x, y)$ is a surface lying above the xy plane.

Joint Probability Distribution

When X and Y are continuous random variables, the **joint density function** $f(x, y)$ is a surface lying above the xy plane.

$P[(X, Y) \in A]$, where A is any region in the xy plane, is equal to the volume of the right cylinder bounded by the base A and the surface.

Joint Probability Distribution

Example 3.9 Suppose that the joint density function is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{elsewhere,} \end{cases}$$

(a) Verify condition 2 of Definition 3.9.

(b) Find $P[(X,Y) \in A]$, where A is the region

$\{(x,y)|0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$

Solution......

Joint Probability Distribution

Example 3.9 Suppose that the joint density function is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere,} \end{cases}$$

(a) Verify condition 2 of Definition 3.9.

(b) Find $P[(X,Y) \in A]$, where A is the region

$\{(x,y) | 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$

Solution......

(b) $P[(X,Y) \in A] = 13/160$

marginal distribution

the discrete case

marginal distribution

the discrete case

the probability distribution $p_X(x)$ of X **alone** is obtained by summing $p(x, y)$ over the values of Y.

marginal distribution

the discrete case

the probability distribution $p_X(x)$ of X **alone** is obtained by summing $p(x,y)$ over the values of Y.

Similarly, the probability distribution $p_Y(y)$ of Y **alone** is obtained by summing $p(x,y)$ over the values of X.

marginal distribution

the discrete case

the probability distribution $p_X(x)$ of X **alone** is obtained by summing $p(x,y)$ over the values of Y.

Similarly, the probability distribution $p_Y(y)$ of Y **alone** is obtained by summing $p(x,y)$ over the values of X.

$p_X(x)$ and $p_Y(y)$: **marginal distributions** of X and Y

marginal distribution

the discrete case

the probability distribution $p_X(x)$ of X **alone** is obtained by summing $p(x, y)$ over the values of Y.

Similarly, the probability distribution $p_Y(y)$ of Y **alone** is obtained by summing $p(x, y)$ over the values of X.

$p_X(x)$ and $p_Y(y)$: **marginal distributions** of X and Y

When X and Y are continuous random variables, summations are replaced by integrals.

marginal distribution

Definition 3.10 The **marginal distribution** of X alone and of Y alone are

$$p_X(x) = \sum_y p(x,y) \quad \text{and} \quad p_Y(y) = \sum_x p(x,y)$$

for the discrete case, and by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

for the continuous case.

marginal distribution

Example 3.10 Show that the column and row totals of Table 3.1 give the marginal distribution of X alone and of Y alone.

marginal distribution

Example 3.10 Show that the column and row totals of Table 3.1 give the marginal distribution of X alone and of Y alone.

TABLE 3.1 Joint Probability Distribution for Example 3.8

$f(x, y)$		$x=0$	$x=1$	$x=2$	Row totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

marginal distribution

Example 3.11 Find **marginal probability density functions** $f_X(x)$ and $f_Y(y)$ for the joint density function of Example 3.9.

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{elsewhere,} \end{cases}$$

marginal distribution

Example 3.11 Find **marginal probability density functions** $f_X(x)$ and $f_Y(y)$ for the joint density function of Example 3.9.

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere,} \end{cases}$$

Solution......

$f_X(x) = \frac{4x+3}{5}$ for $0 \le x \le 1$, and $f_X(x) = 0$ elsewhere.
$f_Y(y) = \frac{2(1+3y)}{5}$ for $0 \le y \le 1$, and $f_Y(y) = 0$ elsewhere.

marginal distribution

The marginal distribution $p_X(x)$ [or $f_X(x)$] and $p_X(y)$ [or $f_Y(y)$] are indeed the probability distribution of the individual variable X and Y, respectively.

marginal distribution

The marginal distribution $p_X(x)$ [or $f_X(x)$] and $p_X(y)$ [or $f_Y(y)$] are indeed the probability distribution of the individual variable X and Y, respectively.

How to verify?

marginal distribution

The marginal distribution $p_X(x)$ [or $f_X(x)$] and $p_X(y)$ [or $f_Y(y)$] are indeed the probability distribution of the individual variable X and Y, respectively.

How to verify?

The conditions of Definition 3.4 [or Definition 3.6] are satisfied.

conditional distribution

recall the definition of conditional probability:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0.$$

conditional distribution

recall the definition of conditional probability:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0.$$

X and Y are discrete random variables, we have

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{p(x,y)}{p_X(x)}, \quad p_X(x) > 0.$$

conditional distribution

Definition 3.11

Let X and Y be two discrete random variables. The **conditional probability mass function** of the random variable Y, given that $X = x$, is

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)}, \qquad p_X(x) > 0.$$

Similarly, the conditional probability mass function of the random variable X, given that $Y = y$, is

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}, \qquad p_Y(y) > 0.$$

conditional distribution

Definition 3.11'

Let X and Y be two continuous random variables. The **conditional probability density function** of the random variable Y, given that $X = x$, is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \qquad f_X(x) > 0.$$

Similarly, the conditional probability density function of the random variable X, given that $Y = y$, is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \qquad f_Y(y) > 0.$$

conditional distribution

Remark:

The function $f(x, y)/f_X(x)$ is strictly a function of y **with** x **fixed**, the function $f(x, y)/f_Y(y)$ is strictly a function of x **with** y **fixed**, both satisfy all the conditions of a probability distribution.

conditional distribution

Find the probability that the random variable X falls between a and b when it is known that $Y = y$

conditional distribution

Find the probability that the random variable X falls between a and b when it is known that $Y = y$

discrete case $P(a < X < b|Y = y) = \sum_{a<x<b} p_{X|Y}(x|y)$

conditional distribution

Find the probability that the random variable X falls between a and b when it is known that $Y = y$

discrete case $\quad P(a < X < b|Y = y) = \sum_{a<x<b} p_{X|Y}(x|y)$

continuous case $\quad P(a < X < b|Y = y) = \int_a^b f_{X|Y}(x|y)dx.$

conditional distribution

Example 3.12 Referring to Example 3.8, find the conditional distribution of X, given that $Y = 1$, and use it to determine $P(X = 0|Y = 1)$.

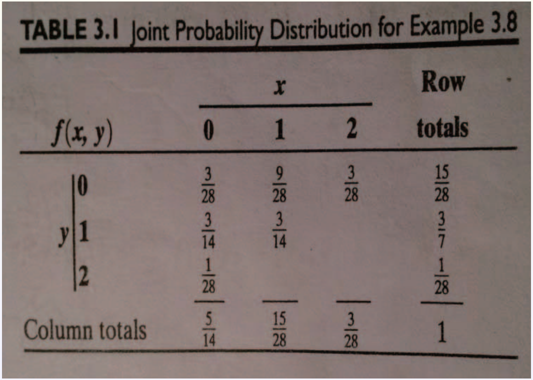

TABLE 3.1 Joint Probability Distribution for Example 3.8

$f(x, y)$		$x = 0$	$x = 1$	$x = 2$	Row totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

conditional distribution

Example 3.12 Referring to Example 3.8, find the conditional distribution of X, given that $Y = 1$, and use it to determine $P(X = 0|Y = 1)$.

TABLE 3.1 Joint Probability Distribution for Example 3.8

$f(x, y)$		$x = 0$	$x = 1$	$x = 2$	Row totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

conditional distribution

Example 3.13 The joint density for the random variables (X, Y) where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces is

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

(a)Find the marginal densities $f_X(x)$, $f_Y(y)$, and the conditional density $f_{Y|X}(y|x)$.
(b)Find the probability that the spectrum shifts more than half of the total observations, given the temperature is increased to 0.25 unit.

conditional distribution

(a)

$$f_X(x) = \frac{10}{3}x(1 - x^3), \quad 0 < x < 1$$

$$f_Y(y) = 5y^4, \quad 0 < y < 1$$

$$f_{Y|X}(y|x) = \frac{3y^2}{1 - x^3}, \quad 0 < x < y < 1$$

(b)

$$P(Y > \frac{1}{2}|X = 0.25) = \frac{8}{9}$$

conditional distribution

Example 3.14 Given the joint density function

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, & 0 < x < 2, 0 < y < 1 \\ 0, & \text{elsewhere}, \end{cases}$$

find $f_X(x)$, $f_Y(y)$, $f_{X|Y}(x|y)$, and evaluate $P(\frac{1}{4} < X < \frac{1}{2}|y = \frac{1}{3})$.

conditional distribution

(a)

$$f_X(x) = \frac{x}{2}, \quad 0 < x < 2$$

$$f_Y(y) = \frac{1+3y^2}{2}, \quad 0 < y < 1$$

$$f_{X|Y}(x|y) = \frac{x}{2}, \quad 0 < x < 2$$

(b)

$$P(\frac{1}{4} < X < \frac{1}{2} | Y = \frac{1}{3}) = \frac{3}{64}$$

statistical independence

events A and B are independent, if

$$P(B \cap A) = P(A)P(B).$$

statistical independence

events A and B are independent, if

$$P(B \cap A) = P(A)P(B).$$

discrete random variables X and Y are independent, if

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all (x, y) within their range.

statistical independence

Definition 3.12 Let X and Y be two discrete random variables, with joint probability distribution $p(x,y)$ and marginal distributions $p_X(x)$ and $p_Y(y)$, respectively. The random variables X and Y are said to be **statistically independent** if and only if

$$p(x,y) = p_X(x)p_Y(y)$$

for all (x,y) within their range.

statistical independence

Definition 3.12' Let X and Y be two continuous random variables, with joint probability distribution $f(x, y)$ and marginal distributions $f_X(x)$ and $f_Y(y)$, respectively. The random variables X and Y are said to be **statistically independent** if and only if

$$f(x, y) = f_X(x) f_Y(y)$$

for all (x, y) within their range.

statistical independence

The continuous random variables of Example 3.14 are statistically independent. However, that is not the case for the Example 3.13.

statistical independence

The continuous random variables of Example 3.14 are statistically independent. However, that is not the case for the Example 3.13.

For discrete variables, requires more thorough investigation.

statistical independence

The continuous random variables of Example 3.14 are statistically independent. However, that is not the case for the Example 3.13.

For discrete variables, requires more thorough investigation. If you find any point (x, y) for which $p(x, y)$ is defined such that $p(x, y) \neq p_X(x)p_Y(y)$, the discrete variables X and Y are not statistically independent.

statistical independence

Example 3.15 Show that the random variables of Example 3.8 are not statistically independent.

TABLE 3.1 Joint Probability Distribution for Example 3.8

$f(x, y)$		x = 0	x = 1	x = 2	Row totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

statistical independence

the case of n random variables

statistical independence

the case of n random variables

The marginal distribution for the discrete case is

$$p_{X_1}(x_1) = \sum_{x_2} \cdots \sum_{x_n} p(x_1, x_2, \ldots, x_n),$$

and the marginal distribution for the continuous case is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) dx_2 dx_3 \cdots dx_n.$$

statistical independence

joint marginal distributions of two r.v. X_1 and X_2

statistical independence

joint marginal distributions of two r.v. X_1 and X_2

discrete case

$$p_{X_1,X_2}(x_1, x_2) = \sum_{x_3} \cdots \sum_{x_n} p(x_1, x_2, \ldots, x_n);$$

statistical independence

joint marginal distributions of two r.v. X_1 and X_2

discrete case

$$p_{X_1,X_2}(x_1, x_2) = \sum_{x_3} \cdots \sum_{x_n} p(x_1, x_2, \ldots, x_n);$$

continuous case

$$f_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) dx_3 dx_4 \ldots dx_n.$$

statistical independence

Definition 3.13 Let $X_1, X_2, \ldots, X_n$ be n discrete random variables, with joint probability distribution $p(x_1, x_2, \ldots, x_n)$ and marginal distributions $p_{X_1}(x_1)$, $p_{X_2}(x_2)$, $\ldots, p_{X_n}(x_n)$, respectively. The random variables $X_1, X_2, \ldots, X_n$ are mutually **statistically independent**,then

$$p(x_1, x_2, \ldots, x_n) = p_{X_1}(x_1) p_{X_2}(x_2) \cdots p_{X_n}(x_n)$$

for all $(x_1, x_2, \ldots, x_n)$ within their range.

statistical independence

Definition 3.13' Let $X_1, X_2, \ldots, X_n$ be n continuous random variables, with joint probability distribution $f(x_1, x_2, \ldots, x_n)$ and marginal distributions $f_{X_1}(x_1)$, $f_{X_2}(x_2)$, $\ldots$, $f_{X_n}(x_n)$, respectively. The random variables $X_1, X_2, \ldots, X_n$ are mutually **statistically independent**, then

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

for all $(x_1, x_2, \ldots, x_n)$ within their range.

statistical independence

Example 3.16 Suppose that the shelf life , in years, of a certain perishable food product packaged in cardboard containers is a random variable whose probability density function is given by

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{elsewhere}, \end{cases}$$

Let X_1, X_2 and X_3 represent the shelf lives for three of these containers selected independently and find

$$P(X_1 < 2, 1 < X_2 < 3, X_3 > 2).$$